08/500844
27 Rec'd PCT/... 04 AUG 1995
PCT/AU94/00150

METHOD AND APPARATUS FOR MANAGING THE STATISTICAL MULTIPLEXING OF DATA IN DIGITAL COMMUNICATION NETWORKS

Field of the Invention

This invention relates to managing the statistical multiplexing of data in digital communication networks which employ the Asynchronous Transfer Mode of multiplexing, transmission and switching.

Background of the Invention

The telephone industry is moving within the Standardisation Sector of the International Telecommunications Union to define the concept of a Broadband Integrated Services Network or B_ISDN. This B_ISDN is to be capable of providing a complete range of services, including telephone, television and data, with all services being multiplexed onto the network over a defined User Network Interface (UNI). Switches within the network are interconnected over a defined Network Node Interface (NNI).

Information for all services within the Broadband Integrated Services Network is conveyed and switched in irregularly occurring fixed size segments called cells, with the method of cell handling referred to as the Asynchronous Transfer Mode (ATM). ATM is a connection oriented technique where ATM level connections are established between service users by means of a call set up or signalling process, or alternatively administratively by network management. Connections may be established for the duration of a call or information transfer, or they may be established on a permanent or semi-permanent basis. The ATM cell has a header block which carries a label necessary to associate cells belonging to the same virtual connection.

In terms of protocol architecture the ATM layer sits immediately above the physical or transmission layer, and has two hierarchical sub-layers. These are the lower Virtual Path (VP) sub-layer and the higher Virtual Channel (VC) sub-layer. The Virtual Channel describes the unidirectional flow of ATM cells marked by a unique identifier called the Virtual Channel Identifier (VCI) which is carried in the cell header. The Virtual Channels are associated by a common unique identifier called the Virtual Path Identifier (VPI) which is also carried in the cell header. Virtual Channels are thus transported within Virtual Paths which are used to aggregate the Virtual Channels.

In accepting connections over the B_ISDN the network operator needs to confirm that sufficient network resources are available to sustain the connection end-to-end at its required Quality of Service and without affecting the service guarantees of the existing connections. To assist in the process of Connection Admission Control the network user is required to characterise its traffic and specify the required Quality of Service. The digital information generated by a source may be characterised by its average bit rate, burstiness, peak duration and other such measures. However for purposes of ATM connection admission, ITU Draft Recommendation I.371 sets out that the traffic on a requested connection be characterised solely by the peak cell rate. If adequate network resources are available to sustain a new connection at its required Quality of Service at the indicated peak cell rate, the connection would be accepted by the network which would then monitor the traffic flow to ensure that the agreed rate is not violated. Although ITU Recommendation I.371 indicates that other parameters besides peak cell rate may be introduced in the future, the present provision is for peak cell rate only, and the provision applies to connections equally at the VP and the VC sublayers.

To assist in the understanding of the significance of the present invention, it is useful to explain why for ATM carried communications in general, definite restrictions on traffic rates over both VP and VC connections are necessary, not only because of the provisions of the ITU Recommendation but for satisfactory operation of the ATM network. An accepted precept for the ATM network is that switching will be done in both the VP and the VC levels. Switching at the VP level would mean that ATM cells entering an input port of a switch would according to the particular VPI label be transferred to a specific output port and be given a specific new VPI label. The VCI portions of the total VPI-VCI label would be unchanged. Switching at the VC level would mean that cells entering an input port of a switch would according to the particular VPI-VCI label be transferred to a specific output port and be given a specific VPI-VCI label.

A switch will in general have multiplicities of input and output ports. The maximum rates at which cells can arrive at input ports and can be taken away from output ports will be set by interface specifications, but must in all cases be definite and finite values. The management of rates at which communications on ATM connections may proceed should ensure that the rate at which cells arrive at any output port of a switch does not exceed the rate at which cells can be taken away from that port.

In the case of VP switching, it must be assured that the sum of rates of the totality of VP connections that are switched to any particular output port does not exceed the capacity of the physical interface at that port. In the case of VC switching it must be assured that the sum of all VC connections that are switched to any particular VP at a particular output port does not exceed the rate allocated to that VP connection, and the sum of rates allocated to VP connections at the given output port does not exceed the

4

capacity of the physical interface at that port. The separate restriction on rates allowed to VP connections that emanate from an output port is made necessary by the possibility that the next switching node would be a VP level switch. Assurances that traffic on VC connections would not exceed set rates would generally be provided by User-Network traffic Contracts and monitoring of that traffic by the network under the provisions of ITU Recommendation I.371 as already indicated. Assurances that traffic on VP connections would not exceed set rates would generally be provided by Connection Admission Control that would not place more VC connections within any given VP connection than can satisfy the given constraint. It will also be possible for a user to obtain a VP connection under contract with the network in an identical manner to that of a VC connection when the assurance that traffic on the VP connection would not exceed that set rate would be obtained in an identical manner to that for a VC connection.

ATM related Quality of Service is to be rated in terms of such parameters as the end-to-end cell transfer delay, cell delay variation, cell loss ratio and cell error ratio. The generally acceptable end-to-end delay will be only of the order of a thousand cell periods. Therefore the allowable delay per switching node can only be of the order of a hundred cell periods (above referred to as cell service intervals). Accordingly, general ATM switches offering either virtual path or virtual channel switching are expected to employ small buffers of no more than several hundred cell capacity with the switch loading kept at such a level as to ensure acceptable cell loss levels resulting from switch buffer overflow. Switch loading is managed by

MWM 06-20-96

the combination of Connection Admission and Route Selection procedures.

S

To provide a data service for N end users over the ATM network one may in the extreme envisage the use of a set of N(N-1) fully meshed connections. These connections would be set up on a permanent or semi-permanent basis reflecting the practice on Local Area Networks (LANs) where no connection set-up is required for sessions. In this respect data networks differ from voice networks where connections are set up and torn down on a regular basis and sessions or calls last only for short periods. If the ATM network, whether private or public, is to offer Quality of Service guarantees those N(N-1) connections across the User Network Interface must be bandwidth resourced with sufficient bandwidth provided to accommodate the peak cell flow rates on these connections. Since data traffic is generally bursty in nature it may be possible with little loss of efficiency to use fewer bandwidth-resourced connections by aggregating the traffic from groups of users by means of statistically multiplexing. Such multiplexing may be at a protocol level above the ATM protocol layer where the information is not yet segmented into ATM cells. This could be at the ATM Adaption Layer (AAL), or even higher.

For data services there are several different AAL layers which may be used and these provide among other functions the segmentation service wherein information for example in the form of data frames from higher layers are mapped into ATM cells and vice versa. However, the multiplexing of whole frames at the AAL layer greatly increases the processing load on the network.

Summary of the Invention

It is an object of the present invention to provide a method and apparatus for managing the statistical multiplexing of data in a digital communication network.

X

According to one aspect of the present invention there is provided a method for supporting a plurality of virtual channel connections within a single virtual path in a digital communications network operating in the Asynchronous Transfer Mode (ATM), where said virtual channel connections have no guarantees of rate at which cells on that connection can be transmitted, but also have no restriction on said rate save that inherent on said virtual path connection, said method comprising the steps of:

storing cells arriving for transmission on said virtual path in a buffer for transmission of cells on said virtual path in conformance with said constraint on said rate;

detecting whether buffer overflow is threatened by the storage of further cells arriving for transmission on said virtual path; and,

while buffer overflow is threatened, admitting for storage in said buffer cells only on such of said virtual channel connections on which the previous cell admitted was not indicated by the header of said previous cell as being the end of transmission on said virtual channel; and,

at all times not admit for storage in said buffer any cells on said virtual channel connections for which since the previous indication of said end of transmission on said virtual channel connection there has been any rejection of cells for storage.

According to another aspect of the present invention there is provided a method for supporting a plurality of virtual paths on a single physical cell transmission system in a digital communications network operating in the Asynchronous Transfer Mode (ATM), each virtual path supporting a plurality of virtual channel connections, where each said virtual path has an individual rate constraint, the method comprising steps of:

storing cells arriving for transmission on any one of said virtual paths in a common buffer for transmission on said physical cell transmission system;

scheduling the stored cells for transmission on said physical cell transmission system at time instants that conform with said individual rate constraints on said virtual paths, and preserving the order of transmission cells on a said virtual path to the order of arrival for said path;

detecting whether buffer overflow is threatened by the storage of further cells for transmission on any of said virtual paths, and detecting whether said scheduling on a particular virtual path is over a specified limit; and for any cell arriving for transmission on a given virtual path,

while buffer overflow is threatened or the scheduling for said virtual path is over the specified limit, admitting for storage in said buffer and scheduling for transmission cells only on such of said virtual channel connections on which the previous cell admitted was not indicated by the header of said previous cell as being the end of transmission on said virtual channel; and,

at all times not admit for storage in said buffer any cells on said virtual channel connections for which since the previous indication of said end of transmission on said virtual channel connection there has been any rejection of cells for storage.

According to another aspect of the present invention there is provided an apparatus for supporting a plurality of virtual channel connections within a single virtual path in a digital communications network operating in the Asynchronous Transfer Mode (ATM), where said virtual channel connections have no guarantees of rate of which cells on that connection can be transmitted, but also have no restriction on said rate save that inherent on said virtual path connection, where said constraint on said

9

virtual path connection is in terms of a specified shortest allowed time interval between successive cells on said virtual path, said apparatus comprising:

means for storing cells arriving for transmission on said virtual path in a buffer for transmission of cells on said virtual path in conformance with said constraint on said rate;

means for detecting whether buffer overflow is threatened by the storage of further cells arriving for transmission on said virtual path; and,

means for admitting for storage in said buffer while buffer overflow is threatened cells only on such of said virtual channel connections on which the previous cell admitted was not indicated by the header of said previous cell as being the end of transmission on said virtual channel, and wherein said admitting means at all times does not admit for storage in said buffer any cells on said virtual channel connections for which since the previous indication of said end of transmission on said virtual channel connection there has been any rejection of cells for storage.

According to another aspect of the present invention there is provided an apparatus for supporting a plurality of virtual paths on a single physical cell transmission system in a digital communications network operating in the Asynchronous Transfer Mode (ATM), each virtual path supporting a plurality of virtual channel connections, where each virtual path has an individual rate constraint said apparatus comprising:

means for storing cells arriving for transmission on any one of said virtual paths in a common buffer for transmission on said physical cell transmission system;

means for scheduling the stored cells for transmission on said physical cell transmission system at time instants that conform with said individual rate constraints on said virtual paths, and preserving the order

9

of transmission of cells on a said virtual path to the order of arrival for said path;

means for detecting whether buffer overflow is threatened by the storage of further cells for transmission on any of said virtual paths, and detecting whether said scheduling on a particular virtual path is over a specified limit; and for any cell arriving for transmission on a given virtual path; and,

means for admitting for storage in said buffer and scheduling for transmission while buffer overflow is threatened or the scheduling for said virtual path is over the specified limit, cells only on such of said virtual channel connections on which the previous cell admitted was not indicated by the header of said previous cell as being said end of transmission on said virtual channel, and wherein said admitting means at all times does not admit for storage in said buffer any cells on said virtual channel connections for which since the previous indication of said end of transmission on said virtual channel connection there has been any rejection of cells for storage.

According to another aspect of the present invention there is provided a method of scheduling the reading out of a data cell from a buffer onto on output comprising the steps of:

calculating a service time S for the reading out of said data cell from said buffer, said service time comprising a delay period measured from the actual time of arrival t_a of a cell to said buffer, said delay equal to at least the difference between an expected arrival time of that cell and the actual arrival time t_a when said cell arrives earlier than expected, and said delay equal to zero when said cell arrives on time or later than expected; and,

reading out said cell at said service time S or at the next time available on said output in the event that

10

said output is not available for reading said cell at said service time S.

According to another aspect of the present invention there is provided an apparatus for scheduling the reading out of data cells from a buffer onto a data output comprising:

means for calculating a service time by reading out said data cells from said buffer, said service time comprising a delay period measured from the actual time of arrival t_a of a cell to said buffer, said delay equal to at least the difference between the expected arrival time of that cell and the actual arrival time t_a when said cell arrives on earlier than expected, and said delay equal to zero when said cell arrives on time or later than expected; and,

means for reading out said cell at said service time S or at the next time available on said output in the event that said output is not available for reading said cell at said service time S.

Brief Description of the Drawings

In order to facilitate a better understanding of the nature of the invention a preferred embodiment of the invention will now be described in detail, by way of example only, with reference to the accompanying drawings in which:

Figure 1 depicts the typical structure of an ATM cell;

Figure 2 depicts the cell header of the ATM cell illustrated in Figure 1 at a User Network Interface;

Figure 3 depicts the cell header of the ATM cell illustrated in Figure 1 at a Network Node Interface;

Figure 4 depicts the B_ISDN layered architecture;

Figure 5 is a representation of a conventional Virtual Path (VP) Switch;

Figure 6 is a representation of a conventional Virtual Channel and Virtual Path switch;

Figure 7 is a representation of an apparatus in accordance with the present invention connected between conventional VP switches;

Figure 8 is a representation of a logical implementation of the buffer and VP rate server of the apparatus represented in Figure 7;

Figure 9 illustrates part of the implementation of an Admission and Write Controller element of the apparatus represented in Figure 8;

Figures 10 and 10A illustrate the remainder of the implementation of the Admission and Write Controller element of the apparatus;

Figure 11 illustrates the implementation of the Schedule part of a Schedule and Read Controller element of the apparatus; and,

Figure 12 illustrates the Read Controller part of the implementation of the Schedule and Read Controller element of the apparatus.

Detailed Description of the Preferred Embodiment

With ATM all information to be transferred across the B_ISDN UNI or NNI is in the form of fixed sized segments called cells. Referring to Figure 1, each cell comprises fifty three (53) octets of information, arranged as a five octet header 10 and a forty eight (48) octet information

field 12. The ATM cell header contains the VPI and VCI fields which carry the labels necessary to associate cells belonging to the same Virtual Path and Virtual Circuit respectively as well as a Payload Type Indicator (PTI).

Table 1 describes the coding recommended for the PTI field in the UNI and NNI cell headers respectively. With this coding bits 4, 3 and 2 refer to the bit position in the ATM cell header 10. When the bit in position 4 is set to zero the cell is characterised as one originating from a network user and sent across the UNI. For such cells the bit in position 2 is available for user signalling. As explained hereinafter, it is required that a cell marking the end of a transmission sequence be identified. Bit 2 can be used to mark a cell which is the last of a transmission sequence.

13

T140X

TABLE 1

PTI Coding Bits 432	Interpretation
000	User data cell, congestion not experienced. ATM-layer-user-to-ATM-layer-user indication = 0
001	User data cell, congestion not experienced. ATM-layer-user-to-ATM-layer-user indication = 1
010	User data cell, congestion experienced. ATM-layer-user-to-ATM-layer-user indication = 0
011	User data cell, congestion experienced. ATM-layer-user-to-ATM-layer-user indication = 1
100	OAM F5 segment associated cell
101	OAM end-to-end associated cell
110	Resource management cell
111	Reserved for future functions

Referring to Figure 2, the structure of the header 10A at the UNI is such that the VPI is contained in the upper four bits and lower four bits of the first and second octets respectively. The VCI is contained in the upper four bits of the second octet, the whole of the third octet and the lower four bits of the fourth octet. The PTI is contained at bit positions 4, 3 and 2 in the fourth octet.

14

The structure of the header 10B at the NNI, shown in Figure 3, differs from that at the UNI in that the whole of the first octet and the lower four bits of the second octet are reserved for the VPI. The location and size of the VCI and PTI fields are the same in the header 10B of the NNI as that in header 10A at the UNI.

The protocol reference model for the B_ISDN is illustrated in Figure 4. Here the lower or physical layer 14 is responsible for the transmission and reception of ATM cells over one of a variety of transmission media and transmission systems. Above the physical layer is the ATM layer 16 which comprises the Virtual Path sublayer 18 and Virtual Channel sublayer 20. At the ATM layer cells from individual VPs and VCs are multiplexed at the UNI into a composite stream for transmission, and arriving cells split into individual tributaries according to their VPI and VCI. The ATM Adaption layer (AAL) 22 sits above the ATM layer and is responsible, among other functions, for the segmentation of large data frames into the ATM cell payloads and vice versa. Finally, higher protocol layers 24 lie above the AAL 22.

Bandwidth resourced Virtual Path and Virtual Channel cell streams, sufficient to accommodate the agreed peak cell rates, are switched within the ATM network. Referring to Figure 5, a VP switch 26 terminates incoming Virtual Paths and cross-connects incoming VPs to outgoing VPs according to the required destinations. The VPI may be reassigned in the switching process but the VCs within each VP remain intact with the VCI values remaining unchanged. In Figure 5 reference numbers 28, 30, 32 and 34 represent physical layer transmission systems which may use, for example, transmission media such as co-axial cables, optical fibres or radio links. Each carry multiple VPs. For example, physical layer 28 carries VPs labelled VP1 and VP2. VP1 is shown being switched to physical layer 30 where it is re-

15

labelled as VP5. VP2 on the other hand is switched to physical layer 34 where it is re-labelled as VP7. In this VP switching process the VCs VC1, VC2, VC3 and VC4 retain their label identifications within their respective VPs.

Figure 6 shows a combined switch comprising a VP switch 36 and a VC switch 38. VC switch 38 terminates both bandwidth resourced VC and VP connections and switches the VCs within a VP independently of each other. VC switch 38 would normally handle bandwidth resourced VCs carrying a mix of traffic types including those like voice and video, which are real-time and delay-sensitive, and accordingly its internal output buffer is relatively small to satisfy the relevant network delay budgets.

In this combined switch VP2 and VP4 are switched only by the VP switch 36 and their VCs remain intact. On the other hand VP1 and VP3 are terminated at the VC switch 38 where the VCs they are carrying are themselves terminated, cross-connected and re-labelled. For example the VC1 and VC2 carried within VP1 are switched through the VC switch 38 into VP7 and VP5 respectively and re-labelled as VC9 and VC11 respectively. Because of switching within the ATM network the VPI and VCI values associated with a particular connection may be different at the two ends, ie at an input UNI and an output UNI.

ATM switches (ie., VP and VC switches), or cross-connecting nodes, are seen from the above descriptions to perform two main functions: the directing of cells from their inputs to their dedicated outputs, and the translation, where relevant, of cell VPIs and VCIs. To avoid cell loss caused by two or more cells simultaneously competing for the same output, internal switch buffers are provided. Because the ATM switches are deployed in switching resourced connections with necessary small throughput delay these

16

buffers would have a capacity of no more than in the order of several hundred cells.

Figure 7 illustrates how an apparatus 56 embodying the present invention would be deployed within an ATM network incorporating an input stage VP switch 42 and an output stage VP switch 46. The apparatus 56 is combined with the cross-connect part of a conventional VC switch 54 to form a Data Switch 40. The Data Switch 40 and a VC switch 44 are connected in parallel between VP switch 42 and VP switch 46.

VP switch 42 switches and terminates incoming bandwidth-resourced VP's on lines 48, some of which are carrying non-bandwidth-resourced VC's and switched onto lines 50 on the output side of VP switch 42. VP's on lines 50 are switched and multiplexed by the Data Switch 40. Remaining lines 52 from VP switch 42 bear VP's which may carry bandwidth-resourced VCs and are switched by the VC switch 44 similar to that described earlier in relation to Figure 6. Buffer overflow in the VC switch 44 may be kept to a very low level by the action of Connection Admission Control exercised by network management. Connection Admission Control limits the number of connections handled and keeps the traffic flow rates on those connections admitted to the rates negotiated at connection set-up.

Apparatus 56 includes output buffers and a VP rate server. The VP rate server controls the writing in and reading out of cells in the output buffers. The output buffers are much larger than those in conventional VC switch 44 which switches bandwidth-resourced VC connections. Transmission lines 58 provide for transmission of cells between outputs of the apparatus 56 and the VP switch 46. Transmission lines 60 provide for transmission of cells between VC switch 44 and VP switch 46. Cells output from VP switch 46 are transmitted on lines 62. It is noted that the cell

flow rates on interconnections 64 between outputs of VC switch 54 and buffer inputs of apparatus 56 are much larger than those on lines 48, 50, 52, 60, 58 and 62. This ensures that the VPs on interconnections 64 can accommodate without loss the possible convergence of concurrent cells from VPs on lines 50.

Notwithstanding the very large output buffers in the apparatus 56, buffer overflow might occur when data bursts occur on several converging streams to an input of the apparatus 56, with consequent cell loss, if it were not for the preventative strategy exercised by the apparatus 56. When buffer overflow is threatened apparatus 56 discards whole newly arriving frames, that is whole VPI-VCI sequences marked with an end of transmission delimiter in the PTI, rather than individual cells from different frames or VPI-VCI sequences. Frame discarding on a cell by cell basis ceases when the buffer occupancy falls below the selected threshold level. In relation to cell output, data cells on the relevant VCs are read from the apparatus 56 buffers by its rate servers on the output VPs on lines 58 according to the rate parameters agreed for each. These VPs, together with those from the VC switch 50 on lines 62, are then able to be switched and multiplexed by the VP switch Output stage 46 in the normal way as described earlier in relation to Figure 5.

Figure 8 shows a logical schematic of one realisation of the apparatus 56. The arrangement of only a single input to one interconnection 64, and a single output on line 58, is shown since this is replicated for the other interconnections 64, and lines 58.

The apparatus 56 includes a buffer 66 for storing cells arriving on an input coupled to interconnection 64, an Admission and Write Controller (AWC) 68 for selectively writing cells into the buffer 66 and a Schedule and Read

Controller (SRC) 70 for selectively reading out cells from the buffer to an output connected to the output line 58. A Delay element 72 is provided to allow sufficient time for cell-header processing by the Admission and Write Controller 68 before each arriving cell is forwarded to a mux 74 for either entry into the Buffer via 76 or discarding via 78.

If the occupancy of the Buffer 66 is such that the Data Switch 40 is accepting new transmissions or data frames, the AWC 68 reads the arriving cells into the appropriate buffer addresses, whether or not these cells are carrying frame segments from the beginning, continuation or end-of-frame, and sends to the SRC 70 on line 84 the address used, the cell VPI and the arrival time t_a for each cell. The AWC 68 also notifies the buffer 66 on line 177 of the address at which the arriving cell is to be stored. If the Buffer occupancy is above the threshold level at which the apparatus 56 does not accept new data frames it reads into the Buffer 66 only those arriving cells which carry the continuation or end-of-frame segments of frames whose first cell had been stored previously, and rejects cells carrying new beginning-of-frame or new frame-continuation and end-of-frame segments. Sufficient storage is kept in the Buffer 66 beyond the level where new frames are rejected to accommodate the continuation of frames that have cells which were already stored in buffer 66 when new-frame discarding commenced. In this way, the AWC acts as a means for detecting whether buffer overflow is threatened by the storage of further cells arriving at its input and admitting for storage in buffer 66 under overload-threat conditions only cells from frames having at least one cell previously stored in the buffer 66.

Figure 9 describes one implementation of the Admit and Write operations of the AWC 68 of the apparatus 56. The modus operandi is such that following the initialisation at

100, the Controller enters a WAIT state 102, where it awaits the arrival of a cell from line 64 onto an input.

When a cell arrives on line 64, the arrival is signalled to the AWC 68 via line 80 and received at 104. At state 106 the AWC reads the cell identifier information such as VPI, VCI and PTI from its header together with its time of arrival, t_a. If the cell is the first to arrive on that particular VPI/VCI, it will not be on an Active-xx list as determined by 108 and the operational state of the AWC 68 proceeds to 110. (The Active_xx lists at 108 include both an Active_Admit list and an Active_Reject list). Then at 112, if the reserve, Res, of memory locations in buffer 66 has fallen below a threshold or limit, the cell VPI/VCI is put in the Active_Reject list and, to avoid possible buffer overflow, the cell is discarded at 114.

The Discard Cell operation at 114 causes a signal to be sent on line 82 (refer Figure 8) to the multiplexer 74, causing it to discard the cell it received via the Delay element 72. This discard operation is in fact a logical rather than a physical one, since Cell Discard causes the cell not to be read into the buffer 66. The Limit used at 110 acts as the decision threshold for the rejection of cells carrying segments from new data frames, ie frames for which no segments had yet been stored in the buffer 66. At 110, if the reserve, Res, of memory address locations has not fallen below the set Limit, and if $t(VPI)-t_a$ is less than the schedule limit, SL, its VPI/VCI is put on the Active_admit list at 116. However, if the cell VPI/VCI read at 108 is on the Active list and not, as tested at 118 marked for discarding by being on the Active-Reject list, the cell is stored by 122 at the current address, C-Addr, fetched by 120 from the address reserve list, Addr_Res. The address C_Addr, together with the cell's time of arrival, t_a, and its VPI are passed by 124 via line 84 to the Schedule and Read Controller, 70. If at 118 it is

20

determined that the cell is on the Active_reject list it is marked for discarding, and subsequently discarded at 126.

When a memory address is taken at 120 from the address reserve list, Addr_Res, the number of addresses in reserve, Res, is decremented by one. Any cell, whether it be written to memory at 122, or discarded at 112 or 126, will after one of these actions eventually have its VPI/VCI removed from the Active_xx list at 128 if it is identified by 130 as containing an end-of-frame segment. End-of-frame information is obtained at 106 from the PTI, held within the cell header 10 (refer figures 1 - 3). The process then returns at 132 to the AWAIT CELL state.

Figure 10 describes one implementation of the Address manager operation of the Admission and Write Controller 68 of apparatus 56. Together, the operations described in Figures 9 and 10 and 10A comprise the whole functionality of the AWC 68. The modus operandi of the address manager is such that following initialisation of the operation at 200 it enters a state, at 202, where it awaits the arrival of a reader signal, R_Sig(C_Addr), from the Schedule and Read Controller 70. The R_Sig (C_Addr) is transmitted on line 86 from the SRC 70 to the AWC 68 (refer Fig. 9).

The availability of the reader signal, R_Sig(C_Addr), is indicated by 204. The reader signal, R_SIG(C_Addr), conveys to the Address manager of the Admission and Write Controller 68 information on which addresses may be added to its address reserve list, Addr_Res, by virtue of the fact that the cells they had contained have been read. The addresses received at 204 are added to the address reserve list, Addr_Res, at 206. Following the addition of an address to the reserve list at 206 the number of addresses in reserve for buffering cells, Res, is incremented by one at 208. The manager then returns at 210 to the AWAIT R_SIG state.

21

Figure 10A describes one implementation of the schedule monitor operation of the Admission and Write Controller 68 of the Data Switch 40. The modus operandi of the schedule monitor is such that following initialisation of the operation at 201 it enters a state, at 203, where it awaits the arrival of a schedule signal, S-Sig(TAT(VPI)), from the Schedule and Read Controller 70 of the Data Switch 40. The S-Sig(TAT(VPI)) is transmitted on line 86 from the SAC 70 to the AWC 68 (refer Fig. 8). The availability of the schedule signal, S-Sig(TAT(VPI)), is indicated by 205. The schedule signal, S-Sig(TAT(VPI)) conveys to the schedule monitor of the Admission and Write Controller 68 the current theoretical arrival time on the particular VPI. This is recorded at 207 as t(VPI). The monitor returns at 209 to the AWAIT S-Sig state.

The output physical link 58 on the output of the apparatus 56 may carry a number of different VPs. The rate of data flow on each of these VPs must be scheduled so as to satisfy the traffic parameters negotiated for each with the ATM network management. The SRC 70 achieves this by the use of rate shaping by virtual scheduling and implemented by linked lists used to read data buffer 66 into bandwidth resourced output VPs. The concept of peak-rate shaping VP connections by not allowing cells on a particular VP to flow with a spacing closer than T, where 1/T is the nominated peak rate, is specified in Recommendation I.371 of the B_ISDN. However, in this embodiment of the apparatus 56 peak rate shaping is implemented by the use of linked lists. The linked list method provides, via a set of pointers, the association and sequence of buffer addresses which contain the ATM cells scheduled for output at each particular cell-start instant on the output VP.

The scheduling of a cell for output at a particular instant involves the calculation of a delay which represents the enforced time spacing between the cell arrival and its

earliest allowable departure time according to the peak-rate scheduling method used. The actual cell departure times will vary from the theoretical due to the unavoidable time-slotting of cell departures and the contention for output service by cells from the various VPs sharing the output line.

In the implementation of the Schedule and Read Controller element 70, VP scheduling is on a first-in-first-out basis according to the negotiated peak cell rate. This follows Recommendation I 371 for the B_ISDN. However VP scheduling is not limited in relation to the nature of the traffic parameter specification used since scheduling may be applied to other regimes such as, for example, average cell rate with restricted burstiness and peak duration.

The VCs within each VP carried on the apparatus 56 output to line 58 are not rate shaped and do not have to be resourced. Switching and multiplexing of these VCs onto a single VP is via the VC switch part 54 of the Data Switch 40 as shown in Figure 7. Scheduling of each VP by the Schedule and Read Controller element 70 of the apparatus 56 and Data Switch 40 is without regard to the VCs they contain.

Figure 11 gives the implementation of the Schedule part of the Schedule and Read Controller 70. The modus operandi of the Schedule is such that following the initialisation of the operation at 220, the Schedule awaits at state 222 the arrival via 84 (refer Fig 8) of the writer signal, W_Sig, from the Admission and Write Controller 68. The availability of the writer signal is indicated by 224 whereby the Schedule receives information that a cell with a particular VPI and which arrived at time t_a has been written to a specific memory address, C_Addr. The time t_a would normally be expressed in units such as microseconds.

23

The expected arrival time of cells with a particular VPI and according to the specific peak rate agreed for that VPI is referred to in the implementation as TAT(VPI). As determined by 226, if the cell arrives later than expected, ie if $TAT(VPI)<t_a$, 228 resets $TAT(VPI)=t_a$. Then for the TAT(VPI) state 230 estimates a scheduling time delay for service, TD equal to the difference between the expected arrival time TAT(VPI) and the actual arrival time t_a, ie

$$TD=TAT(VPI)-t_a$$

Hence if a cell arrives earlier than expected the schedule programs a delay for service, TD.

At 232 the expected arrival time for the next cell with that particular VPI is calculated based on the last value of TAT(VPI) plus T, where T is the reciprocal of the peak cell transmission rate for that VPI. Thus

$$T(VPI)=1/(\text{Peak cell rate for VPI})$$

At 233 the new valve of TAT(VPI) is sent as the signal S-Sig(TAT(VPI)) on line 86 to the admission and write controller 68.

At 234 TD is expressed in terms of an integer number of cell service intervals, ND, where the cell service interval is the reciprocal of the cell transmission rate at the output to line 58, ie

$$ND = \text{Ceiling}(TD/\text{cell transmission rate at } 58)$$

wherein Ceiling (x) is x if x is a whole integer; and, the next whole integer greater than x if x is not a whole integer.

24

At 236 the cell is scheduled for service at time S, expressed as ND service intervals after the current service interval beginning at time N. The range of service periods, R, is sufficiently large that a single VPI buffer could fill before R would be exhausted. The apparatus 56 contains a buffer for each VPI. The individual buffers are ideally logical divisions within a single physical buffer. At 238 the scheduled service time for the cell which arrived at time t_a and which is stored at address C-Addr is appended to a linked list to be operated on by the Read Controller part of the Schedule and Read Controller element 70. The process then returns at 240 to the AWAIT W_Sig state.

Figure 12 gives the implementation of the Read Controller part of the Schedule and Read Controller 70. The Read Controller reads onto the output connected to line 58 cells which have been scheduled for service by the Schedule. Timing for the read operation, ie Cell Start, is derived internally from the digital network clock at 58. The clock pulse coincident with the beginning of the first octet of a cell forms a Cell Start. On the initialisation of the process at 250 an AWAIT state is entered at 252 where it awaits the arrival of the Cell Start. After a Cell Start arrives at 254, the Read Controller at state 256 fetches the linked list of tasks scheduled for service at that time. If, as determined at state 258, no tasks are scheduled for the current cell service interval, ie if there are no cells to be read, 260 increments N by one and operation at state 262 awaits the next Cell Start. If, as determined at state 258, there are tasks scheduled for the current cell service interval, N, the Read Controller at state 264 removes the first entry from the linked list at N and 256 then reads to the output the cell at the appended address. Since this address location is thus freed state 268 advises the Admission and Write Controller via the R_Sig on line 86. If, as determined in state 270, there

25

are no other tasks scheduled for N, state 260 increments N by one and operation awaits the next cell start at state 262. Otherwise the remaining tasks are re-scheduled by 272 to the next cell period where they will be served ahead of tasks already scheduled at N+1.

The implementations given in Figures 9, 10, 10A, 11 and 12 do not describe how the mutual exclusion of shared resources, such as variables and linked lists, is achieved. Techniques for this are well known to persons skilled in the art and contained in references such as by M.Ben-Ari, ("Principles of Concurrent and Distributed Programming" Prenice Hall, 1990).

By the application of the preferred embodiment of this invention, particularly in the form of a Data Switch in an ATM network, it is possible to discard whole VPI-VCI sequences or whole frames at the ATM protocol layer and avoid multiple frame destruction through cell loss multiplication by discarding of individual cells from multiple frames when buffer overflow is threatened. Such a management strategy, if implemented in the ATM adaption layer or higher layers would greatly increase processing overheads.

As is evident from the above description the apparatus 56 and Data Switch 40 enables switching and multiplexing of the data streams carried on non-bandwidth resourced, non rate shaped VC connections within bandwidth resourced, rate shaped VPs.

Notwithstanding the large buffers in the apparatus 56 there may still be congestion when data bursts occur on several VC data streams converging to a single input. The apparatus 56 is however able to operate under sustained overload conditions when ATM cells arrive at its output buffer 66 at a rate higher than the VP service rate. Under

26

these conditions, when buffer overflow is threatened, the apparatus 56 discards whole frames rather than individual cells originating from different frames as would be the case with normal ATM cell multiplexing. The apparatus 56 is able to identify which cells mark the beginning and end of the segmented data frame by means of identifier information in the cell heading such as a Payload Type Indicator (PTI) which delineates the end of frame on a particular VC. When buffer overflow occurs in the conventional ATM-layer cell multiplexer, single discarded cells can result in the loss of whole segmented data frames comprising multiple cells producing cell loss multiplication.

Since embodiments of the present invention incorporate very large buffers, with a correspondingly long fill-time, it is possible to reactively renegotiate with the ATM network management the level of bandwidth resource associated with the serving VP. Consequently, when a buffer is seen to be filling rapidly, one may increase the server VP bandwidth providing for a higher cell, and hence frame, output rate. Such reactive negotiation of network resources is less effective with conventionally sized ATM switch buffers and networks which cover wide areas with very high speed links.

Various prior art schemes have been suggested to counter buffer overflow in ATM switches but these differ from that used by the Data Switch which also avoids cell loss multiplication. One such scheme, described by Gersht and Lett, ("A Congestion Control Framework for ATM Networks" IEEE Journal on Selected Areas in Communications, Vol. 9, No. 7, September 1991, pp 1119-1130), includes the use of feedback control signals to limit or choke for a period the cell flow rate on VCs feeding the overflowing buffer. Another scheme, described by Turner, ("Design of Local ATM Networks" INFOCOM 1993), proposes the dropping of a fraction of the tributary VCs. This has the unfortunate

effect of stopping all data traffic on those stopped VCs. On the other hand, in the Data Switch of the preferred embodiment, cells associated with whole new frames are discarded while guaranteeing the passage of those frames which are in transit. No particular VC is cut off and each has an equal chance of getting a data frame through the network when the buffer occupancy falls sufficiently to allow readmission of new frames.

Various cell flow shaping methods, to achieve particular traffic parameters, may be applied to the VC cell flows to the apparatus 56. The implementation of the preferred apparatus 56 described herein uses, by way of example and not by way of limitation, unshaped VC cell flows reflecting the practice on Local Area Networks (LANs) where no bandwidth restrictions are placed on data input rates other than those set by the capacity of the physical layer transmission. Within the apparatus implementation described the bandwidth restriction is set by that allowed for the VPs carrying the VC data flows.

Numerous variations and modifications to the described embodiment of the apparatus and method of the invention will suggest themselves to persons skilled in the electronics and the telecommunications art, in addition to those already mentioned, without departing from the basic inventive concepts. For example, another method that may be used as a basis for admitting or discarding each newly arrived frame is one which makes a decision based on the current value of two parameters, viz the current level of buffer fill, and the number of frames which have been accepted by the Data Switch and which are currently in transit through it. In this context, a frame "in transit" is one which has at least one of its cells in the Data Switch buffer awaiting output service. With this method a new frame would be accepted if it is judged that the level of unused buffer capacity is sufficient to cater for

28

transit as well as the cells from the new frame. Furthermore, the apparatus of the invention does not preclude a mode of operation wherein it can handle within the one implementation the switching and multiplexing of both bandwidth resourced and non-bandwidth resourced VCIs although these two categories of VCI must necessarily be associated with different VPIs. This is necessary since the non-resourced VCIs would not be switched by the normal VCI switches whereas the bandwidth resourced VCIs may be. In its preferred form, the invention is implemented in an ATM digital communications network, however it may also have application in other environments. All such variations and modifications are to be considered within the scope of the present invention, the nature of which is to be determined from the foregoing description and the appended claims.

24